Indian Journal of Experimental Biology
Vol. 54, January 2016, pp. 26-36

Cloning and characterization of a water deficit stress responsive transcription factor gene from *Oryza sativa* L.

Ibandalin Mawlong, Kishwar Ali & Aruna Tyagi*

Division of Biochemistry, Indian Agricultural Research Institute (IARI), New Delhi 110 012, India

Received 12 June 2014; Revised 01 October 2014

Understanding the biochemical and molecular basis of drought mechanism in rice is important as drought is one of the major causes affecting rice crop adversely. A 1017 bp gene sequence encoding AP2/ERF family TF was isolated from *Oryza sativa* sp. Indica *cv* N22 encoding a protein of 338 amino acid residues, with a molecular weight of 36.58 kDa, and no intron in the ORF. The gene was named as *AP2/ERF*-N22(2) different from the drought responsive gene *AP2/ERF*-N22 that we reported earlier. *AP2/ERF*-N22(2) has entirely different characteristics from that of *AP2/ERF*-N22. It has a single AP2 domain of 55 amino acid residues and a cluster of acidic amino acid residues at the C-terminal region, which could function as a trans-activation domain. Presence of NLS indicates that it is a nuclear localized transcription factor encoding gene. It falls in group VI L, sharing characteristic similarities. Arabidopsis members of group VI L have been shown to be involved in response to cytokinin under drought stress.

Keywords: AP2-Apetela type 2 transcription factor, Drought, Ethylene responsive factor (ERF), N22, Paddy, Rice crop

Paddy is the most important cereal crop in East Asia. It requires comparatively more water. However, availability of water for irrigation purpose in countries like India is a matter of concern. To develop crop plants with enhanced tolerance to drought stress, a basic understanding of physiological, biochemical, and gene regulatory networks is essential. Response to environmental stresses occurs at all levels of organization. Transcription factors (TF) play pivotal role in signal transduction to activate or suppress defense gene expression, as well as in the regulation of interaction between different signaling pathways possibly by regulating expression of downstream genes as *trans*-acting elements via specific binding to *cis*-acting elements in the promoters of target genes. TFs, as opposed to most structural genes, tend to control multiple pathways, and thereby emerged as powerful tools for manipulation of complex metabolic pathways in plants. Among the different TFs, the ethylene responsive transcription factors (ERF) family plays significant role in plant growth, enabling plants capable of fighting ambiance changes[1,2]. Therefore, it is important to understand the function of these genes in order to boost crop yield.

*Correspondence:
Phone: +91 11 25842038
E-mail: at_bio@iari.res.in

The AP2/ERF family was first identified in *Arabidopsis*[3] followed by other crops such as grapevine[4] poplar[5] and rice[6] with 145, 132, 200, and 170 genes, respectively. The members of the AP2/ERF family can be divided into 3 groups based on their overall structure[7]. Members of the AP2 subfamily (14 in *Arabidopsis*) contain two AP2/ERF domains; members of the RAV subfamily (6) contain an AP2/ERF domain and an additional B3 DNA-binding domain, while the other members (125) contain only a single AP2/ERF domain[7,8]. Sakuma *et al.*[8] analyzed the phylogenetic relationships of the 125 single-AP2/ERF members in *Arabidopsis* based on the similarity of their AP2/ERF domain. These 125 single-AP2/ERF members are further classified into 3 groups: the DREB subfamily (56 members; group A), the ERF subfamily (65 members; group B) and others (4 members). The DREB and ERF subfamilies are both divided into 6 subgroups, A-1 to A-6 and B-1 to B-6, respectively. DREB1/CBFs belong to subgroup A-1, and DREB2s belong to subgroup A-2. Three of four 'other' members are similar to AP2 subfamily members and seem to have a typical second AP2/ERF domain, while the fourth 'other' protein (AT4G13040) has a unique type of AP2/ERF domain. Nakano *et al*.[9] reported a molecular, phylogenetic and motif analysis of AP2/ERF members in *Arabidopsis* and rice. Their results are similar to the classification of Sakuma's[8].

But Nakano *et al*.[9] also found that the A-1 and A-4 subgroups and the A-2 and A-3 subgroups share common conserved motifs, respectively, suggesting their common origin.

Overexpression of AP2/ERF transcription factor gene in transgenic plants such as rice[10], tobacco[11], *Arabidopsis*[12], oilseed rape[13] resulted in enhanced tolerance to drought[14], salt[15] and freezing[16]. However, recently Zhuang *et al*.[17] showed contradictory results where a rice OsAP23, functioning as a AP2/ERF transcription factor, upon overexpression reduces salt tolerance and rate of germination. This complicates the characterization of AP2/ERF genes, which require in depth study of each gene within the given classification. Thus, elucidating the function of transcription factor genes would pave way to crop improvement.

In light of this, earlier we had reported isolation and characterization of an *AP2/ERF*-N22 from rice[1]. In the present study, we did isolation, cloning and molecular characterization of another such stress responsive AP2/ERF family TF encoding gene from rice (*Oryza sativa* L. *cv* N22).

Material and Methods

Plant material and drought stress treatment

In the present study the seeds of *Oryza sativa* L. drought tolerant *cv* N-22, an early maturing, deep rooted, adapted to upland conditions and drought susceptible *cv* IR 64, were used. All plants were grown in growth chamber of National Phytotron Facility, IARI, New Delhi, in soil mixture, which comprised one part of sand and perlite and two parts of compost. Seeds were sown at the density of 6 plants per 4-cm pot, and after one week seedlings were transplanted in 16-cm pot with a density of 4 plants per pot. Nutrients (Hoagland solution) were supplied 10 days after germination. On 45th day after transplanting, the plants were subjected to water deficit stress (WDS) by withholding water. The relative water content (RWC) of leaf was estimated according to method of Barrs and Weatherley[18]. Leaves of control and drought stress-treated plants were harvested, immediately dropped in liquid nitrogen and stored at −80°C prior to RNA extraction.

DNA Isolation and PCR amplification

The isolation of genomic DNA was done using CTAB method by Rogers and Bendich[19] from etiolated rice seedlings. The DNA was purified using CsCl/ethidium bromide equilibrium ultracentrifugation as described in the manual[20]. About 10 ng of genomic DNA was used for PCR. The sequence of forward and reverse primers were as follows:

AP2/ERF-N22(2) Forward ATG GCC CCT TCC AAG CAG CAG CA
AP2/ERF-N22(2) Reverse TCA CAC GCA GAA ATC CAC CTC CC

The reaction conditions for PCR included a denaturing step of 95°C for 3 min, followed by 35 cycles of 30s at 94°C, 30s at 55°C, and 1 min at 72°C, ending with an elongation step of 5 min at 72°C. Genomic amplicon was separated on a 1.0% (w/v) agarose gel. PCR amplified fragment was gel eluted and cloned into the T/A cloning vector, pGEM^{T}-Easy from Promega according to the manufacturer's protocol. Recombinant plasmids were transformed using the competent cells of *E. coli* DH5α as host as described by Sambrook *et al*.[20]. Blue white screening was used to select the recombinants. The plasmid DNA was isolated from host *E. coli* DH5α by alkaline lysis mini-preparation method. The recombinant plasmid isolated was restricted with *Eco*RI enzyme to check the presence of the insert.

Semi-quantitative expression analysis

Isolation of total RNA from the leaves by Tri-reagent was done according to manufacturer's protocol. Samples were collected for RNA extraction at 3, 6 and 9 days after withholding water and control seedlings. Total RNA of control and drought stressed leaf tissue was run on 1.0% agarose gel containing 0.05 μg/ml ethidium bromide and 20 mM guanidium thiocyanate (GTC). RT-PCR involves two steps, the first strand cDNA synthesis and PCR using gene specific primers. cDNA synthesis was done using RevertAid H Minus First Strand cDNA synthess kit (Thermo Scientific) following manufacturer's protocols. Expression analysis was done by semi quantitative RT PCR.

DNA sequencing and *in silico* analysis

The clone was sequenced. The nucleotide and amino acid sequences were compared with those released in GenBank databases by using the BLAST program. Translation of the nucleotide sequence and primary structural analysis of the predicted protein were done on ExPASy Server (http://expasy.org/tools/dna.html). CLUSTAL W and BLAST are available at URL site www.genome.ad.jp and BioEdit software.

Results and Discussion

The AP2/ERF family is a large group, associated with not only biotic and abiotic stresses but also various growth and developmental functions. Thus, to

characterize their function has had relevance. The primers were designed for a putative transcription factor gene (LOC_Os01g04020) from TIGR rice database of a specific AP2/ERF transcription factor gene based upon the presence of ABRE and DRE/CRT *cis* elements in promoter region (600 bp upstream of initiation codon ATG) of *Oryza sativa* sp. Japonica, a closest relative of *Oryza sativa* sp. Indica.

Genomic DNA isolated from Indica *cv* N-22 was amplified. The amplicon size of 1017 bp was obtained (Fig. 1). The amplicon was checked on agarose gel, sliced, eluted and cloned in the T/A cloning plasmid vector, pGEMT-Easy. The recombinant plasmids were identified by blue and white screening and recombinant plasmid DNA was isolated. The restriction analysis of plasmid DNA was done by single digestion using *Eco*RI (Fig. 2). The cloned gene was named as AP2/ERF-N22(2).

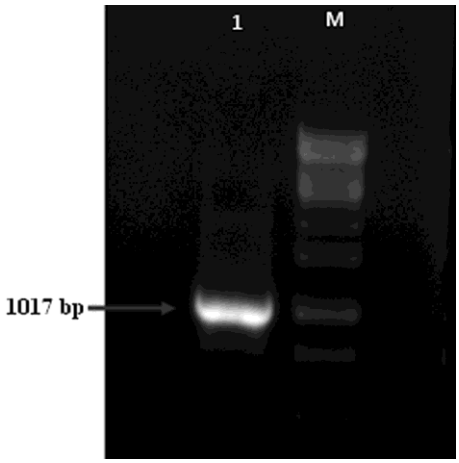

Fig. 1— Agarose gel showing PCR amplified gene from rice *cv* N22. Lane 1, 1017 b Amplicon; and lane M, 1 kb ladder.

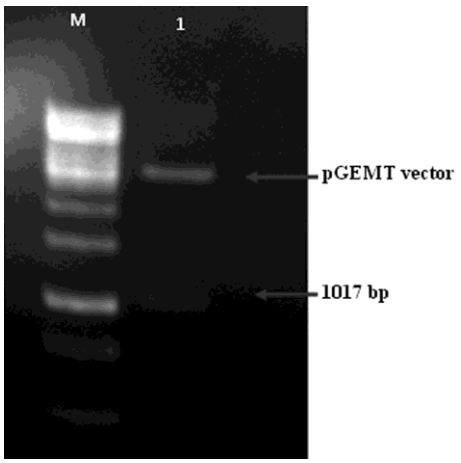

Fig. 2— Agarose gel showing plasmid DNA restricted with *Eco*RI. Lane 1, 1017 bp restricted product; and Lane M, 1kb gene ladder.

RT PCR analysis of AP2/ERF N22 expression

Semi-quantitative RT PCR was done to confirm the induction of gene under water deficit stress conditions. Two cultivars namely, drought tolerant N22 and drought susceptible IR-64 were chosen for expression study. Actin was used as a quantitative control. Expression analysis of *AP2/ERF*-N22(2) was done at different stress levels i.e. 3, 6 and 9 days after withholding water. Expression analysis at different stress levels is shown in Fig. 3. In N22, induction of *AP2/ERF*-N22(2) was observed to be more as compared to IR-64. In N-22, expression increased with the increase in water deficit stress (WDS) i.e., number of days after withholding water. Maximum expression was observed at 9 days after withholding water. In case of IR-64, no change in its expression with respect to control was observed at 3 and 6 days after withholding water, and induction of the gene was observed only under severe stress condition i.e., after 9 days of withholding water. This confirms the induction of gene under WDS conditions (Fig. 3).

***In silico* analysis of the gene**

The insert had a sequence of 1017 bp. ORF finder predicted a cDNA without any intron. Complete nucleotide sequence amplified is shown in Fig. 4. The gene sequence was submitted to Gene Bank with Accession # KC988330.1. NCBI Blast N showed 100% homology to many clones from *O.sativa* sp. Japonica group (Accession no. EU837259.1, NM_001048461.1, AY341828.1,AK119645.1, AP002526.1) having AP2 binding domain, but none belonged to Indica species, which confirmed the isolated gene *AP2/ ERF*- N22 (2), possibly first report from Indica sp. Nucleotide composition analysis showed that the gene has 62% (G+C) and 38% (A+T) content. This verifies that it is a gene having an AP2 binding domain.

Protein sequence analysis

The *AP2/ERF*-N22(2) encodes a protein of 338 amino acid residues, molecular weight of 36.5 kDa

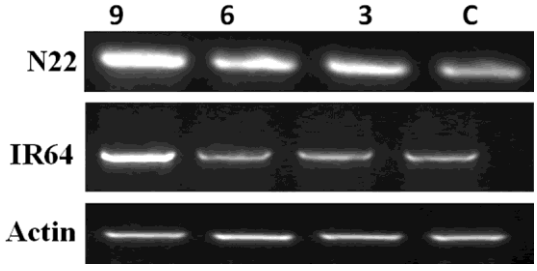

Fig. 3— Expression analysis of *AP2/ERF*-N22(2) in N22 and IR64 under control (C) and water deficit stress condition. Numbers (3-9) represents the days after withholding water.

and theoretical pI of 4.75. The predicted protein contains 53% of negatively charged amino acid residues (Glu and Asp) and 38% of positively charged amino acid residues (Arg and Lys). NCBI BlastP homology search revealed 100% homology to *O. sativa* sp. Japonica of unknown function, suggesting *AP2/ERF*-N22(2) is a novel gene and its function has not yet been annotated.

In the putative protein sequence three repeat units were found using www.ebi.ac.uk/tools/pfa/radar/ Analysis tool. The first repeat starts from sequence 221 to 239 which was found repeated twice at positions 240-255 and 256-271. The second repeat sequence starts from 173-185 which was repeated only once at position 191-203. The third repeat sequence was found at position 285-299 which was repeated only once at position 301- 318 (Fig. 5). According to Magnani *et al.*[21], horizontal transfer of an HNH-AP2 endonuclease from bacteria or viruses into plants may have led to the origin of the AP2/ERF family of transcription factors via transposition and homing processes.

Nakano *et al.*[9] did a classification of AP2/ERF family in both rice and *Arabidopsis*, and classified AP2/ERF family of rice into 14 groups. Nakano *et al.* grouped *Arabidopsis* (At1g25470, At1g 49120, At1g68550, At3g25890) and rice [LOC_Os09g13940.1 (accession #NP_001062835.1), LOC_Os08g27220.1 (accession #BAD03821.1), LOC_Os01g04020.1 (accession #KC988330.1), accession #NP_001041926.1, accession #EAY72401.1] into VI L. It is assumed *AP2/ERF*-N22(2) which is homologous to LOC_Os01g04020 also belongs to VI L and had all the characteristic features of group VI L. The characteristic features of group VI L, according to Nakano *et al.*[9], are absence of introns in the coding region; an imperfect AP2 domain, molecular mass of 31-36 kDa, length of amino acids between 330 and 338 residues, presence of conserved motifs CMV, conserved putative MAP kinase; and/or casein kinase I phosphorylation sites.

Functional domain analysis

Generally, different regions in transcription factors contain functionally important domains involved in transcriptional activity, protein-protein interactions, DNA binding and nuclear localization[22]. Proteins within a subgroup that share these motifs are likely to share similar functions.

ATGGCCCCTTCCAAGCAGCAGATGCTCCTCAAGAAGGTGATGGCGAAGAAGCCCAAG
ACGAAGAGGCTATCCGGGTTTGGCCTTAAACCCTCTGCTGCTTTCTCCAGGCCCCATGTG
CCGTCGGCGGCGGCCTCCCTGCAGCCAAGCCGTCGCGTCCGCGTCGTGTTCGAGGACCCT
GACGCGACGGACTCCGACTCTGACGACGACGAGGACGCCGGCGCCGCCTCAAAGAAGCGC
TACTTTGAGCTCTTCATTGGCAAGCCAGCTTCGTCGACGAAGCAGGCCTCTCCGGCGTCC
ACCGTCGCTGCCTACGCCAACATCGGCAAGGTTGGGAGCACCTGCTACCGTGGTGTGCGC
CTCCGGAAGTGGGGCAAGTGGGCGGCGGAGATCCGCAACCCCTTCACCGGCCATAGGGAG
TGGCTTGGCACCTTTGACACTGCTGATGCGGCCTCCGCCGCCTACCAGTCTGCCTCCCGC
AACTTCGCCGAAGAGAAGCGCCGCCGTGGTGTGGCCTCCTCTGCCTCACCTGCTAGTTCG
GCGACCCCTACTCCGACTGCGTCCTCGTCGTCCTCGACTTCTGCCGCGCCATTCGCCCAC
CCTTCGCCGTCGTCTGTGCTCGAAGCCACCAAGCCAGCTCCAAAGCCAGAGTCGCCGCCG
CTGCCGGAGCAAGCTGCAACTCCCCTCCTGGTGGAGGCTACCAACGAGACCGCCGAGCTG
CCGGATGACCCAGAGTTCTACAAGGATATACTGCGCGGTCTACAGCTGCCGGACATTGAC
CCGATGGATTTCCGAGCTGGGCTGGATGCTCTGGATATCTCCGATGTGCCGGCTTACATG
AATGGCGAACAAGACGTACTCTTCACTGAGGACATGCTGCTTGGAGACTTCGCTGAAGAA
GATGACCTCGACCTCGACGACATCGGTGATGACTTCTGCGAGGATTTCCCAGAGATACCC
AGCGGCTACGACTTCGGCCGTGGTGATATGTT**CCGGCAGGTGGATTTCTGCGTGTGA**

Fig. 4— The complete nucleotide sequence of the gene obtained after sequencing the amplicon. Bold letters represent forward and reverse primers.

MAPSKQQQMLLKKVMAKKPKTKRLSGFGLKPSAAFSRPHVPSAAASLQPSRRVRVVFEDP
DATDSDSDDDEDAGAASKKRYFELFIGKPASSTKQASPASTVAAYANIGKVGSTCYRGVR
LRKWGKWAAEIRNPFTGHREWLGTFDTADAASAAYQSASRNFAEEKRRRGVASSASPASS
ATPTPTASSSSTSAAPFAHPSPSSVLEATKPAPKPESPPLPEQAATPLLVEATNETAEL
PDDPEFYKDILRGLQLPDIDPMDFRAGLDALDISDVPAYMNGEQDVLFTEDMLLGDFAEE
DDLDLDDIGDDFCEDFPEIPSGYDFGRGDMFRQVDFCV

Fig. 5— Presence of repeated sequences in *AP2/ERF*-N22(2) as predicted at www.ebi.ac.uk/Tools/pfa/radar/. Putative protein sequence showing three repeats in different colours.

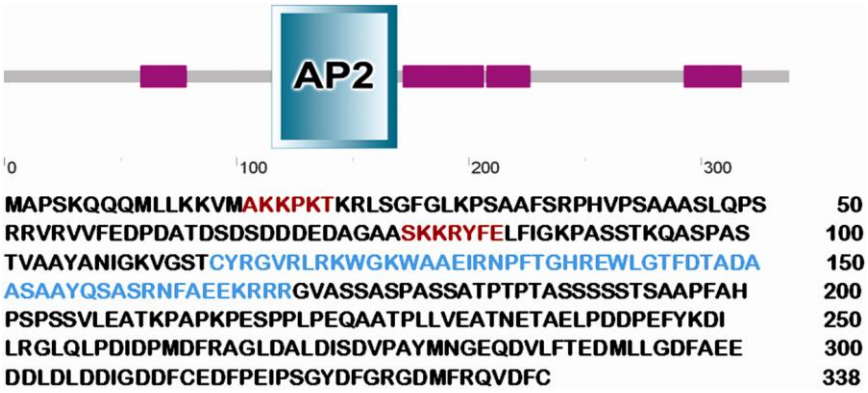

Fig. 6— SMART domain search shows the presence of single AP2 domain starting from 115 to 169 (55) amino acid residues. The predicted nuclear localized sequences are from position 16 to 21 and from 77 to 83.

Analysis of the sequence using SMART domain search (http://smart.embl-heidelberg.de/), found the putative gene to contain a single AP2 DNA domain, which starts form 115th and ends at 169th amino acid (55 amino acid residues) (Fig. 6). This domain binds to the GCC/DRE region. Generally, most of the AP2 domains range from 60-70 amino acid residues[23]. However, AP2 domain in *AP2/ERF*-N22(2) was found to be imperfect as it was of 55 amino acid residues long. The imperfect AP2 domain classifies this gene into group VI L. AP2 domain is highly conserved. The presence of the YRG and RAYD amino acid sequences is a typical characteristic of AP2/ERF domain. In *AP2/ERF*- N22(2) two regions of amino acid residues YRG and RAYD were found to be conserved in AP2 domain (Fig. 7).

The YRG region (YRG element) is of about 20 amino acid long N-terminal stretch rich in basic and hydrophilic residues. It was proposed to have a role in the DNA binding by making a direct contact with the DNA because of its basic character[24]. The RAYD region of about 40 amino acids (RAYD element) in its C-terminal sequence has a domain of 18 amino acids capable of forming an amphipathic α-helix and is thought to have an important role in function of the domain[24,25]. The RAYD element was proposed to mediate protein-protein interactions through α-helix or to have an alternative role in DNA binding through interactions of hydrophobic face of the α-helix with the major groove of DNA.

ClustalW2 alignment of group VI L members of rice with *AP2/ERF*-N22(2) sequence showed AP2 domain, with highly conserved YRG and RAYD element in all sequences (Fig. 8). The YRG element corresponds to the N terminal loop, β strand 1 and β strand 2. The RAYD element corresponds to βstrand 3, the alpha helix and C terminal loop. In this study, AP2 domain had 55 amino acid residues, with YRG element of 18 amino acid residues and the RAYD region consisting of 37 amino acid residues out of which 18 amino acid residues were capable of forming an amphipathic α-helix. In most of AP2 domains, 14th Ala and 19th Asp are conserved[8]. But in this study, only the 14th Ala was conserved, and 19th amino acid Asp, was found to be substituted by Asn. In all group VI L sequences, Asn was found to be conserved at the 19th position.

Nuclear localization sequence

Nuclear Localization Signal (NLS) is an amino acid sequence which directs a protein for import into the cell nucleus by nuclear transport. Typically, the signal consists of one or more short sequences of positively charged Lys or Arg exposed on the protein surface. The NLSs are classified into two classes i.e., classical NLS and non-classical NLS. Classical NLSs can be further classified as either monopartite or bipartite. ClustalW2 of all selected sequences showed maximum basic amino acid residues concentrated at the N terminal region, which could act as NLS (Box 2 & 3 in Fig.7). The predicted protein has a bipartite NLS of the sequence AKKPKT (starting from 16 to 21) and SKKRYFE (starting from 77 to 83) (Fig. 6) as predicted in ELM, the data base of eukaryotic linear motifs (elm.eu.org). Amino acid sequence SKKRYFE was shown to be conserved in all members belonging to group VI L. Chelsky *et al*.[26] proposed the consensus sequence K-K/R-X-K/R for monopartite NLSs. The NLS of nucleoplasmin, KR [PAATKKAGQA] KKKK, is the prototype of the ubiquitous bipartite signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids[27]. Both signals are recognized by importin α.

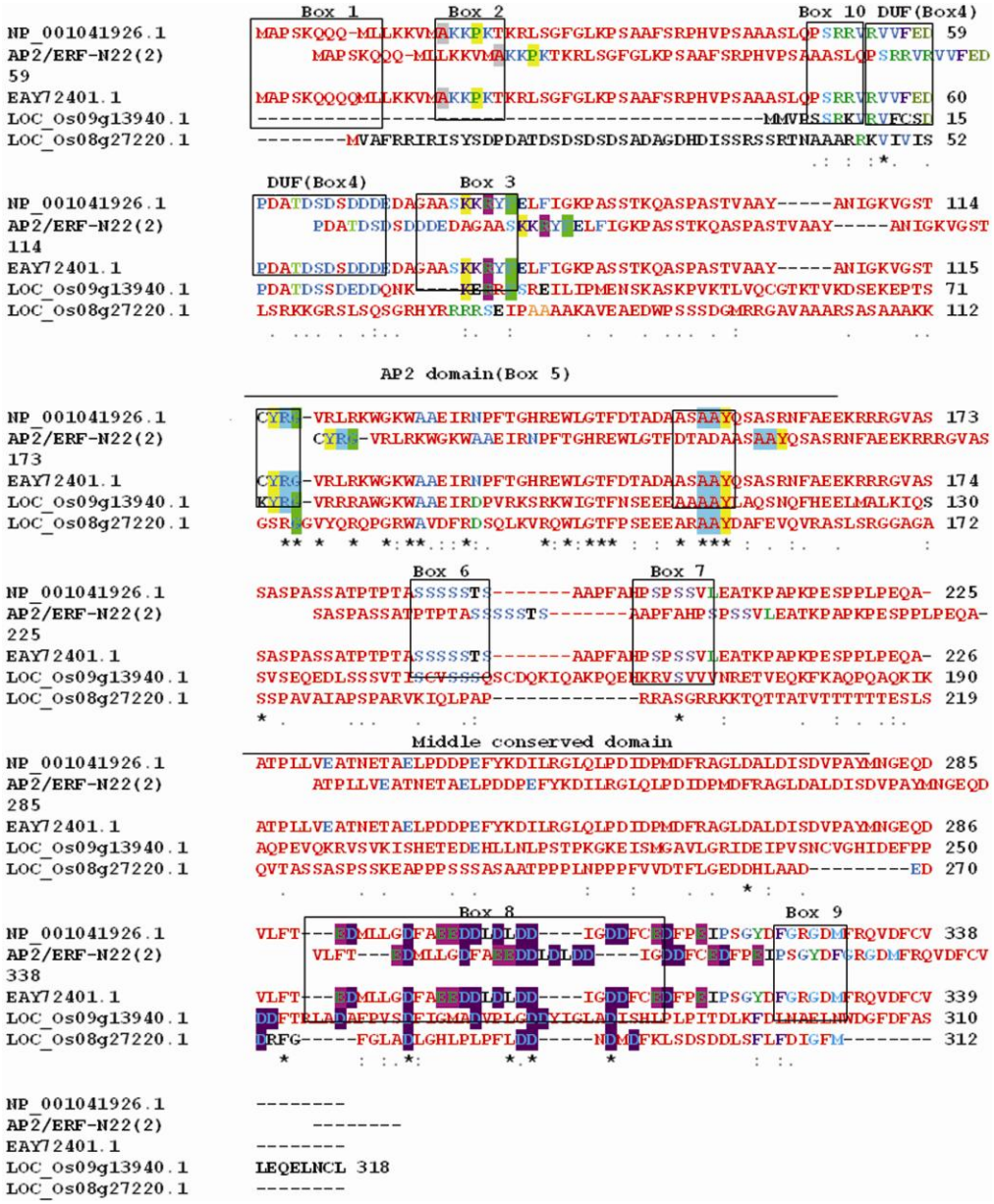

Fig. 7— ClustalW2 multiple alignment of sequences belonging to *Oryza sativa* group VI L. [Box 1 represents highly conserved N-terminal domain; Box 2 and 3 represent bipartite NLS starting from 16 to 21 and another starting from 77 to 83; Box 4 represents the DUF; Box 5 represents the AP2 domain of 55 amino acid residues, within this domain there are two separate regions showing the YRG and RAYD elements; Box 6 represents conserved region rich in Ser and Thr (starting from 188 to 194); Box 7 represents the MAP kinase region (202 to 207); Box 8 represents the conserved region rich in Asp (D) and Glu (E) residues, conserved at 290th to 318th position; Box 9 represents the conserved C- terminal domain; and Box 10, Arg rich region (50 to 55)].

The importin α contains a bipartite NLS itself, which is specifically recognized by importin β.

There are many other types of NLS, such as the acidic M9 domain of hnRNP A1, the sequence KIPIK in yeast transcription repressor Matα2, and the complex signals of snRNPs. Most of these NLSs appear to be recognized directly by specific receptors of the importin β family without the intervention of an importin α-like protein[28]. Lee *et al.*[29] recently classified another type of NLSs called PY-NLSs, so named because of the proline-tyrosine amino acid pairing in it, allowing the protein to bind to importin β2 (also known as transportin or karyopherin β2), which then translocates the cargo protein into the nucleus. Krizek and Sulli[30] had also identified the sequence KKKR for nuclear localization of protein in AINTEGUMENTA

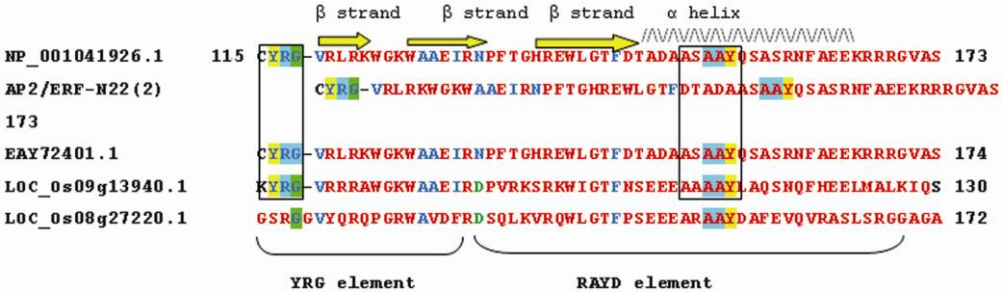

Fig. 8— ClustalW2 alignment of groupVI L showing conserved AP2 domain in all selected sequences. [Strand 1, starting from 119th V-122th R; Strand 2, starting from 126th W- 133thN; Strand 3, starting from 138th H- 146th D; and Alpha helix, starting from 147th T- 164th E. The YRG element corresponds to the N terminal loop and strands 1 and 2. The RAYD element corresponds to strand 3 and the alpha helix and c terminal loop]

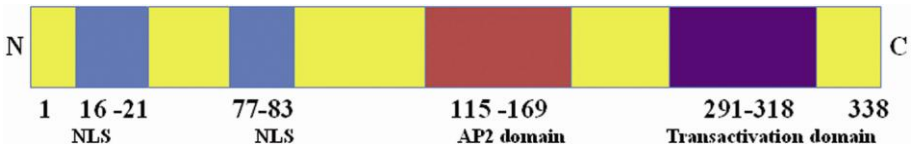

Fig. 9— Overview of different domains present in of AP2/ERF-N22(2) protein.

protein in Arabidopsis, which is a member of AP2/ERF family. Tsi1 (for Tobacco stress-induced gene1) encoding an EREBP/AP2–type transcription factor that enhances resistance against pathogen attack and osmotic stress in Tobacco, belonging to group VI[9] *also* had bipartite NLS at the N terminal region[11] similar to *AP2/ERF*- N22(2). The analysis of the NLS suggested it to be nuclear localized transcription factor.

Transcriptional activation domain

Functional domain analysis of *AP2/ERF*- N22(2) showed an additional domain outside the DNA binding domain which according to Liu *et al.*[31], is involved in transcriptional activity and protein-protein interactions. ClustalW2 alignment of genes from the group VIL showed maximum acidic amino acids Asp and Glu clustered from 290th position to 318th position at the C-terminal domain that could act as transactivation domain (Box 8 in Fig. 7). Regions rich in acidic amino acid residues are often designated as transcriptional domain[31]. Transcriptional activation domains (TAD) are regions of transcription factor which in conjugation with a DNA binding domain can activate transcription from a promoter by contacting transcriptional machinery either directly or through other proteins known as co-activators. Based upon the amino acid composition, it can be classified into acidic domain, glutamine-rich domain or proline-rich domain. *AP2/ERF*- N22(2) has an acidic transcription activation domain.

Domain of unknown function

AP2/ERF-N22(2) showed another highly conserved domain of unknown function starting from 56 to 71 amino acids (Box 4 in Fig. 7). One DUF named ESK1 functions as a negative regulator of cold acclimation. Mutation in DUF protein, *ESKIMO 1* resulted in enhanced freezing tolerance of Arabidopsis[32], which implies their importance in stress response. Bischoff *et al.*[33] in their studies showed another DUF protein called TBL involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers. The DUF sequence present in *AP2/ERF*-N22(2) suggests that it has an important role not only to withstand stress but also in plant growth and development. The overview structure of all the functional domains of an *AP2/ERF*- N22(2) is shown in Fig. 9.

Multiple sequence alignment

ClustalW alignment was done for *AP2/ERF*-N22(2) with other sequences from *Oryza sativa* belonging to group VI L [LOC_Os09g13940.1 (accession #NP_001062835.1), LOC_Os08g27220.1 (accession #BAD03821.1), LOC_Os01g04020.1 (accession #KC988330.1, accession# NP_001041926.1 and accession# EAY72401.1). An investigation of the conserved motifs in the proteins of *AP2/ERF*-N22(2) with proteins of group VI L in the ERF family of rice was carried out via multiple alignment analysis with ClustalW2 *www.ebi.ac.uk/Tools/msa/clustalw2*.

Result from the alignment of selected sequences from group VI L, showed a cluster of Arg (50-55 amino acid residues) (Box 10 in Fig. 7), cluster of serine and Thr (188 to 194) residues (Box 6 in Fig. 7) and conserved SPSSVL (starting from 202 to 206). Clusters of Ser and Thr and serine–rich SPSSVL (Box 7 in Fig. 7) could be involved in phoshorylation. Clusters of Ser–rich residues were found in several AP2/ERF proteins including ERFs and might be involved in activation of transcription[23]. According to Nakano *et al*.[9], SPSSVL sequence complies with putative MAP kinase and/or casein kinase I phosphorylation sites conserved in group VI and VI-L proteins. Closer investigation of *AP2/ERF*-N22(2) showed a conserved middle domain with all the members of the Group VI L (Fig. 7), a characteristic feature of group VI-L.

Motif scan analysis of AP2/ERF-N22(2) protein available at the site http://myhits.isb-sib.ch/cgi-bin/motif_scans showed the following sites: N-glycosylation site (NETA) at positions 235-238, cAMP- and cGMP-dependent protein kinase phosphorylation site (KRLS) at positions 22-25, casein kinase II phosphorylation site (TDSD,SDD,SVLE,SGYD) at positions 63-66, 67-70, 205-208, 321-324, N-myristoylation site (GLKPSA, GTFTDA,GVASSA) at positions 28-33, 143-148, 170-175 and protein kinase C phosphorylation site (TYR,SRR,SKK,STK) at positions 21-23, 50-52, 77-79, 92-94 which could impart important role in activating the protein. A consensus nucleotide sequence, AGCCGCC known as the GCC-box, was first identified in the promoter region of the pathogenesis-related genes. Binding to this element mediates responses to different abiotic stresses[34]. Phosphorylation of Os*EREBP1* by *BWMK1* results in the enhancement of its binding to GCC box and the trans activation activity mediated transcription[35]. Transcription factors DREB1A/CBF3 and DREB2A specifically interact with *cis*-acting dehydration-responsive element/C-repeat (DRE/CRT) involved in cold and drought stress-responsive gene expression in *Arabidopsis thaliana*. Intact DREB2A expression does not activate downstream genes under normal growth conditions, suggesting that DREB2A requires posttranslational modification for activation[36]. DREB2A transcription factor controls water deficit-inducible gene expression and requires post-translational modification for its activation.

Pyre2 software shows the presence of three anti parallel β strands and one α helix strand. The deduced

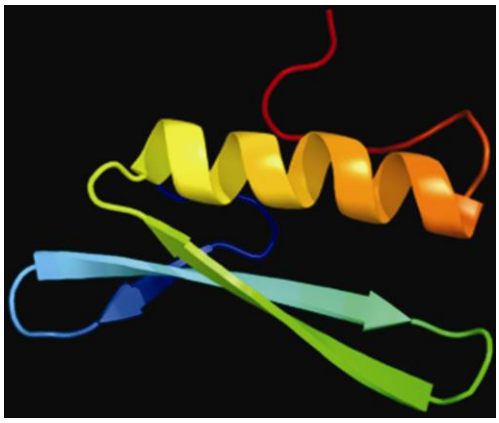

Fig. 10— Secondary structure of AP2/ERF-N22(2) protein.

3D model of the AP2 domain as predicted by pyre2 software is shown in Fig. 10. Secondary structure determined by Pyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/webscripts) software reveals binding of AP2 domain to GCC box, which is present in most promoter region of genes responsive to drought and cold[34]. Secondary structure of the gene *AP2/ERF*-N22(2) predicts the presence of three anti-parallel β sheets, comprising of strand 1 (119 Val-122 Arg), strand 2 (126 Tyr-133 Arg) and strand 3 (138 His-146 Asp) packed along α helix (147 Thr-164 Glu) (Fig. 10). The target DNA bends slightly at the central CG step, thereby allowing the DNA to follow the curvature of the β sheets[37].

Closer observation of AP2/ERF-N22(2) showed the presence of two Ala residues at 128 and 129 positions in the α helix, and hydrophobic amino acids residues Val 119, Phe 147 and Ile 131 in the β-strand. ClustalW2 alignment showed these amino acid residues to be highly conserved among members of group VI L (Fig. 7). The geometry of α helix relative to the β-sheets appears to be determined by the interaction of many Ala residues and the larger hydrophobic residues in the β sheets[37]. According to Allen *et al*.[37], there are four hydrophobic amino acid residues in the β-sheets which clamp the α helix at the four corners. But in case of AP2/ERF-N22(2), there are only three amino acid residues, within the β-sheets.

TAIR BlastP of all the Arabidopsis members of group VI L suggested their role in cytokinin response. Since AP2/ERF-N22(2) fulfills all the characteristic features mentioned for group VI L such as absence of introns, acidic pI, almost identical molecular weight and length of amino acid residues, conserved SPSSVL rich region, imperfect AP2 domain and

conserved CMV domain. This strongly suggests that AP2/ERF-N22(2) might have similar kind of functions as group VI L proteins. It is possible that *AP2/ERF*-N22(2) could have a possible role in cyokinin response.

The plant hormone cytokinin is recognized as an essential regulator of plant root systems. It is known that cytokinin inhibits root elongation and branching[38]. Cytokinin (CK) levels tend to decrease under adverse environmental conditions. A general view has emerged that during stress; a reduction of CK supply from the root alters gene expression in the shoot, and thereby elicits appropriate responses to ameliorate the effects of stress. However, studies have indicated that transcription of many stress-inducible genes can also be induced by CK application.

A correlation between root system and resistance to water stress has been found in several crop plants, and breeding attempts have focused on obtaining cultivars with larger root systems[39]. For example, drought-resistant rice (*Oryza sativa*) varieties have a deeper and more highly branched root system than drought-sensitive varieties[40]. Plants with larger root systems have an increased ability to compete for nutrients and survive under water deficit condition. Recently, it was shown that a reduction of cytokinin status in plants causes the formation of a larger root system. Different approaches have been used to reduce the cytokinin status, including constitutive overexpression of cytokinin degrading cytokinin oxidase/dehydrogenase (CKX) genes[41], reduction of cytokinin biosynthesis[42], mutation of cytokinin receptor genes[43], or suppression of the cytokinin signaling pathway[44]. It was shown that a reduced cytokinin status retards the exit of dividing cells from the root meristem, thus increasing the number of cells in the meristem[43,45]. Genetic analyses have shown that cytokinin acts at the transition zone via the transcription factor IAA3/SHY21. Transgenic *Arabidopsis thaliana* and tobacco (*Nicotiana tabacum*) plants with enhanced root-specific degradation of the hormone cytokinin, a negative regulator of root growth causes enhanced root growth, drought tolerance, and leaf mineral enrichment in *Arabidopsis* and tobacco[46].

Broad spectrum of genes are expressed on exposure to dehydration. Thus, understanding the function of these genes is far from complete. The gene *AP2/ERF*-N22(2) is novel and first to be reported from *Oryza sativa* sp. Indica *cv* N22. Expression studies showed *AP2ERF*–N22(2) to be drought inducible transcription factor gene. Since group VI L Arabidopsis members have been assigned a role in cytokinin response, manipulations of cytokinin, a hormone which negatively regulates root growth as stated by various reports, may be important and an alternative strategy against drought stress. It was hypothesized that *AP2/ERF*-N22(2) could have functions related to cytokinin response. Its functions could be explored by using transgenics, and may have significant role under water deficit stress. The hypothesis derived from this study needs to be tested in model plants to understand its role in drought and the interaction and molecular relationship of genes within the network of stress genes.

References

1 Mawlong I, Ali K, Kurup D, Yadav S & Tyagi A, Isolation and Characterization of an AP2/ERF type drought stress inducible transcription factor encoding gene from rice. *J Plant Biochem Biotechnol*, 23 (2014) 42.

2 Mawlong I, Ali K, Srinivasan R, Rai R D & Tyagi A, Functional validation of a drought-responsive AP2/ERF family transcription factor-encoding gene in Arabidopsis. *Mol Breeding*, 35 (2015) DOI 10.1007/s11032-015-0290-9.

3 Zhou J, Tang X & Martin GB, The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *EMBO J*, 16 (1997) 3207.

4 Jaillon OAJM, Noel B & Policriti A, The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449 (2007) 463.

5 Zhuang JCB, Peng RH, Zhu B & Jin XF, Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. *Biochem Biophys Res Comm*, 371 (2008) 468.

6 Rashid M, Guangyuan H, Guangxiao Y, Hussain J & Xu Y, AP2/ERF Transcription factor in rice: Genome-Wide canvas and syntenic relationships between monocots and dicots. *Evol Bioinformatics*, 8 (2012) 321.

7 Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK & Yu GL, Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 290 (2000) 2105.

8 Sakuma YLQ, Dubouzet JG, Abe H, Shinozaki K & Yamaguchi-Shinozaki K, DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. *Biochem Biophys Res Comm*, 90 (2002) 998.

9 Nakano T, Suzuki K, Fujimura T & Shinshi H, Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiol*, 140 (2006) 411.

10 Ishizaki T, Maruyama K, Obara M, Fukutani A, Yamaguchi-Shinozaki K, Ito Y & Kumashiro T, Expression of *Arabidopsis* DREB1C improves survival, growth, and yield of upland new rice for Africa (NERICA) under drought. *Mol Breeding*, 31 (2013) 255.

11 Park JM, Park CJ, Lee SB, Ham BK, Shin R & Paek KH, Overexpression of the tobacco *Tsi1* gene encoding an EREBP/AP2–Type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. *Plant Cell*, 13 (2001) 1035.

12 Xiong AS, Jiang HH, Zhuang J, Peng RH, Jin XF, Zhu B, Wang F, Zhang J & Yao QH, Expression and function of a modified AP2/ERF transcription factor from *Brassica napus* enhances cold tolerance in transgenic *Arabidopsis*. *Mol Biotech*, 53 (2013) 198.

13 Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T & Thomashow MF, Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol*, 127 (2001) 910.

14 Bihani P, Char B & Bhargava S, Transgenic expression of sorghum DREB2 in rice improves tolerance and yield under water limitation. *J Agri Sci*, 149 (2011) 95.

15 Golldack D, Lüking I & Yang O, Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Reports*, 30 (2011) 1383.

16 Maki HSM, Ogawa K, Kaida R, Yamamoto N & Kaneko ST, Cloning and expression profile of an ERF gene isolated from cold-stressed poplar cells (*Populus nigra*). *Cytologia*, 76 (2011) 11.

17 Zhuang J, Jiang HH, Wang F, Peng RH, Yao QH & Xiong AS, A Rice *OsAP23* functioning as an AP2/ERF transcription factor, reduces salt tolerance in transgenic *Arabidopsis*. *Plant Mol Biol Rep*, 31 (2013) 1336.

18 Barrs HD & Weatherley PE, A re-examination of the relative turgidity technique for estimating water deficits in leaves. *Aus J Biol Sci*, 15 (1962) 428.

19 Rogers SO & Bendich AJ, Extraction of DNA from plant, fungal and algal tissue, In: *Plant Molecular Biology Manual*, (Eds. SB Gelvin & RA Schilperoort, Kluwer Academic Publishers, Boston, MA), 1994, 1.

20 Sambrook J, Fritsch EF & Maniates T, Laboratory Manuals (2^{nd} edn)(Cold Spring Habor Laboratory Press, New York), 1998.

21 Magnani E, Lander KS & Hakea S, From endonucleases to transcription factors: Evolution of the AP2 DNA binding domain in plants. *Plant Cell*, 16 (2004) 2265.

22 Liu L, White MJ & MacRae TH, Transcription factors and their genes in higher plants functional domains, evolution and regulation. *Eur J Biochem*, 262 (1999) 247.

23 Riechmann JL & Meyerowitz EM, The AP2/EREBP family of plant transcription factors. *Biol Chem*, 379 (1998) 633.

24 Okamuro JK, Caster B, Villarroel R, Montagu MV & Jofuku KD, The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc Nat Acad Sci*, 94 (1997) 7076.

25 Jofuku KD, Boer BGW, Montagu MV & Okamuro JK, Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell*, 6 (1994) 1211.

26 Chelsky D, Ralph R & Jonak G, Sequence requirements for synthetic peptide-mediated translocation to the nucleus. *Mol Cell Biol*, 9 (1989) 2487.

27 Dingwall C, Robbins J, Dilworth SM, Roberts B & Richardson WD, The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. *J Cell Biol*, 107 (1988) 841.

28 Mattaj I W & Englmeier L, Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem*, 67 (1998) 265.

29 Lee BJ, Cansizoglu AE, Süel KE, Louis TH, Zhang Z & Chook YM, Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell*, 126 (2006) 543.

30 Krizek BA & Sulli C, Mapping sequences required for nuclear localization and the transcriptional activation function of the *Arabidopsis* protein AINTEGUMENTA. *Planta*, 224 (2006) 612.

31 Liu QKM, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K & Shinozaki K, Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought-and low-temperature-responsive gene expression, respectively in *Arabidopsis*. *Plant Cell*, 10 (1998) 1391.

32 Xin ZG & Browse J, *Eskimo*1 mutants of Arabidopsis are constitutively freezing-tolerant. *Proc Natl Acad Sci*, 95 (1998) 7799.

33 Bischoff V, Nita S, Neumetzler L, Schindelasch D, Aurelie U, Ravit E, Staffan P, Deborah D & Wolf-Ru¨diger S, Trichome Birefingence and its homolog *AT5G01360* encode plant-specific DUF231 proteins required for cellulose biosynthesis in *Arabidopsis*. *Plant Physiol*, 153 (2010) 90.

34 Ohme-Takagi M & Shinshi H, Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell*, 7 (1995) 173.

35 Cheong YH, Moon BC, Kim JK, Kim CY, Kim MC, Kim IH, Park CY, Kim JC, Park BO, Koo SC, BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. *Plant Physiol*, 132 (2003) 1961.

36 Sakuma Y, Maruyama K, Osakabe Y, Qin F, Seki M, Shinozaki K & Yamaguchi-Shinozaki K, Functional analysis of an *Arabidopsis* transcription factor,DREB2A, involved in drought-responsive gene expression. *The Plant Cell*, 18 (2006) 1292.

37 Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M & Suzuki M, A novel mode of DNA recognition by a b-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *The EMBO J,* 17 (1998) 5484.

38 Werner T & Schmulling T, Cytokinin action in plant development. *Curr Opin Plant Biol*, 12 (2009) 527.

39 Price AH, Cairns JE, Horton P, Jones HG & Griffiths H, Linking drought-resistance mechanisms to drought avoidance in upland rice using a QTL approach: Progress and new opportunities to integrate stomatal and mesophyll responses. *J Exp Bot*, 53 (2002) 989.

40 Price AH, Tomos AD & Virk DS, Genetic dissection of root growth in rice (*Oryza sativa* L.) I: A hydroponic screen. *Theor Appl Genet*, 95 (1997) 132.

41 Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H & Schmulling T, Cytokinin-deficient transgenic Arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell*, 15 (2003) 2532.

42 Miyawaki K, Matsumoto-Kitano M & Kakimoto T, Expression of cytokinin biosynthetic isopentenyltransferase

genes in *Arabidopsis*: Tissue specificity and regulation by auxin, cytokinin, and nitrate. *Plant J*, 37 (2004) 128.

43 Riefler M, Novak O, Strnad M & Schmulling T, *Arabidopsis* cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *Plant Cell*, 18 (2006) 40.

44 Heyl A, Ramireddy E, Brenner WG, Riefler M, Allemeersch J & Schmulling T, The transcriptional repressor ARR1-SRDX suppresses pleiotropic cytokinin activities in *Arabidopsis*. *Plant Physiol*, 147 (2008) 1380.

45 Dello Ioio R, Linhares FS, Scacchi E, Casamitjana-Martinez E, Heidstra R, Costantino P & Sabatini S, Cytokinins determine *Arabidopsis* root-meristem size by controlling cell differentiation. *Curr Biol*, 17 (2007) 678.

46 Werner T, Nehnevajova E, llmer I K, Nova´Ondrˇej K, Strnad M, Kra¨mer U & llinga TS, Root-Specific Reduction of Cytokinin causes Enhanced Root Growth, Drought Tolerance, and Leaf Mineral Enrichment in *Arabidopsis* and Tobacco. *Plant Cell*, 22 (2010) 3905.